Comparison Of High Dynamic Range Near-Neighbor Detection Approaches For TPF[1,2]

Gene Serabyn
Jet Propulsion Laboratory
MS 171-113
4800 Oak Grove Drive
Pasadena, CA 91109
818-393-5243
gene.serabyn@jpl.nasa.gov

Abstract— While observations of stellar Doppler shifts have recently uncovered a population of massive planets orbiting nearby stars, the direct detection of radiation from non-solar planets is still beyond reach, because of the high contrast ratios and small angular separations involved. However, a number of high-contrast near-neighbor detection approaches have the potential to enable the direct detection of planets as faint as the Earth from space-borne platforms. Several concepts, ranging from large-aperture visible-wavelength coronagraphs to multiple-spacecraft infrared interferometers, are currently being considered in the context of NASA's Terrestrial Planet Finder program. Because of their demanding requirements, identification of the most feasible near-term approach calls for careful consideration of the options. This paper compares the optical requirements inherent in two of the primary candidate architectures: coronagraphy and nulling interferometry. At the same wavelength, both approaches share similar phase control requirements, but nulling has the advantage that it is sensitive closer to the optical axis. On the other hand, infrared nulling eases phase matching requirements, but introduces the need for longer baselines and cryogenic cooling.

TABLE OF CONTENTS

1. INTRODUCTION
2. LARGE APERTURES AND APODIZATION
3. SCATTERED LIGHT
4. DIMMING THE CENTRAL STAR
5. CORONAGRAPHY
6. NULLING INTERFEROMETRY
7. COMPARISON SUMMARY

1. INTRODUCTION

While evidence for planets around nearby stars continues to accumulate [1], the direct detection of light from such extra-solar planets, either reflected or thermal, remains beyond present capabilities, due to the small (few arcseconds or less) angular separations and huge contrast ratios (millions to billions) involved. In fact, both of the detection techniques that have been successfully applied to date (stellar radial velocity variations and the dimming of the star during a planetary transit) rely on indirect observations, i.e., observations of the parent stars rather than of the planets themselves.

In order to directly detect very faint planets in close proximity to vastly brighter stars, the limitations of current observational capabilities must be overcome. In practice, either telescope diameters need to be increased to the point at which the much fainter light from planets can emerge from the glare of the central star's diffraction pattern, or the scattered and diffracted light in the inner region of the focal plane needs to be strongly suppressed by a combination of aperture apodization and wavefront error reduction, or the star must be dimmed relative to its surroundings by employing one of a number of starlight rejection techniques, such as coronagraphy or nulling interferometry.

Although all of these approaches have the potential to provide higher dynamic range detection capabilities, they also all require substantial development before successful deployment on a space mission can be considered. As a result, NASA is currently conducting studies to determine which of these technologies might be the most viable and cost effective in the near term, for application to the proposed Terrestrial Planet Finder (TPF) mission or a precursor thereto. Given that the options are very diverse, ranging from large-aperture visible-wavelength coronagraphs to infrared interferometers involving separated cryogenic spacecraft, the selection of the approach to be used for sarlight suppression has tremendous implications for the architecture and cost of the TPF mission. The goal of this patper is to provide a brief overview and comparison of the classes of starlight suppression techniques currently being considered, in order to examine the nature of the optical tradeoffs involved. The paper begins with a brief review of the simplest cases – large

[1] 0-7803-7231-X/01/$10.00/© 2002 IEEE
[2] IEEEAC paper #245, Updated December 14, 2001

apertures and apodized apertures - and then moves on to a discussion of the coronagraphic and interferometric approaches. Finally, a comparison of these concepts is provided in the last section.

2. Large Apertures and Apodization

Conceptually, the most obvious route to faint near-neighbor detection is finer angular resolution brought about by increased telescope diameters. The resultant narrowing of the diffraction-limited core of the focal-plane point spread function (PSF) then yields a larger star-planet separation in terms of the angular resolution, λ/D. (With λ the observing wavelength, and D the aperture diameter, λ/D is roughly the full width at half maximum of the resultant focal plane Airy pattern). In principle, shorter wavelengths are also advantageous, but this resolution gain is counteracted by the increasing importance of small wavefront (i.e. surface) errors at shorter wavelengths, which translates into an increasingly bright halo of scattered light around the PSF's diffraction core. This effectively sets a lower limit to the observing wavelength. Adaptive optics (AO) wavefront correction can be used to reduce the scattered light halo, but this places increasingly stringent demands on the AO system as the wavelength is decreased. Thus, the observing wavelength cannot be decreased arbitrarily.

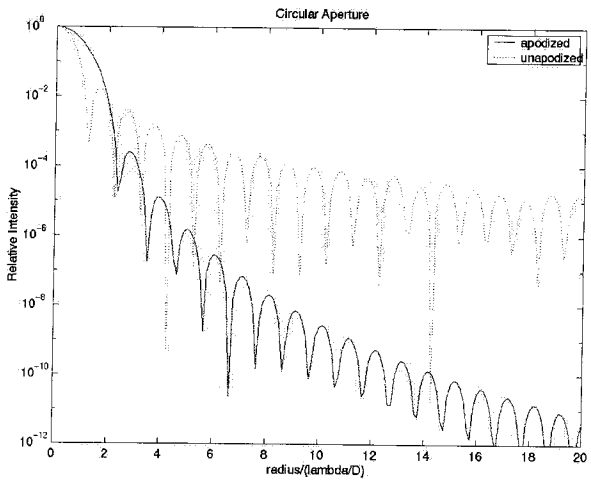

Figure 1. Comparison of ideal diffraction patterns for ideal apodized and unapodized circular apertures. The apodization function used was $(1-r^2)^3$, where r is the normalized radius. The off-axis distance is given in units of the angular resolution of λ/D.

Taking λ = 0.5 μm as the minimum wavelength under consideration, λ/D for representative 1 and 10 m apertures is then 0.1″ and 0.01″, respectively. In comparison, 0.1″ is the angular offset that the Earth's 1 AU (Astronomical Unit) orbit about the Sun would subtend if seen from a distance of 10 pc, while 1″ corresponds to Saturn's 10 AU orbital radius seen from that distance. Thus, given the contrast ratio of order 10^{-10} between the Earth and the Sun at visible wavelengths, much larger telescope diameters are required to resolve faint, close-in planets from their parent stars. Considering diffraction from an ideal circular aperture, the intensity in the PSF wings drops to the level of 10^{-9} to 10^{-10} of the central value only by the time an off-axis angle, θ, of several hundred λ/D is reached. This results from the fact that the successive sidelobe peaks in the perfect diffraction pattern only fall off approximately as θ^{-3} (e.g., Figure 1).

As is well known, the diffraction sidelobes can be greatly suppressed by smoothly tapering or ``apodizing'' the aperture. Indeed, with an appropriate choice of apodization function, the off-axis intensity can be suppressed to a remarkable level (Figure 1), reducing sidelobes to the 10^{-10} level beyond ≈ 10 λ/D, or 1″ (0.1″) off-axis for a 1 (10) m aperture. The price to be paid for this gain is decreased angular resolution (Figure 1) and throughput. Alternatively, diffraction sidelobes can be suppressed over selected regions of the focal plane simply by opting for non-circular apertures [2,3]. For example, the familiar example of the square aperture [2] has prominent sidelobes extending primarily along two orthogonal axes perpendicular to the aperture's edges (with sidelobe peaks decreasing approximately as θ^{-2}), leaving the diagonal regions of the diffraction pattern comparatively clear (with diffraction peaks falling off as θ^{-4}.)

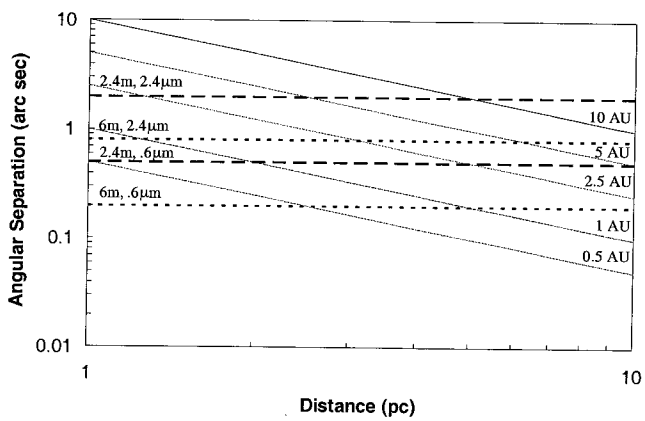

Figure 2. Solid curves: angular separations for fixed orbital radii vs. stellar distance. For comparison, the dashed and dotted curves show the 10 λ/D angular radii for 2.4 m diameter (HST; dashed) and 6 m diameter (NGST; dotted) telescopes at representative wavelengths of 0.6 μm and 2.4 μm.

If diffraction sidelobes can be so easily suppressed, then why haven't such techniques already been applied on moderate to large telescope apertures? Certainly for the stars located within several parsecs of the Sun, the technique would seem especially promising (Figure 2), at least for space-based telescopes, where atmospheric turbulence is not an issue. Figure 2 shows the angular separations subtended by potential planets at a selection of orbital radii, as a function of

the stellar distance, and compares these to the $10\lambda/D$ angular radii for two aperture diameters: 2.4 m and 6 m (representative of the Hubble Space Telescope [HST] and the Next Generation Space Telescope [NGST]), at wavelengths of 0.6 and 2.4 μm). For example, with a telescope of the diameter of the HST, and a wavelength of 0.6 μm, $10\ \lambda/D = 0.5''$. A Jupiter analog located 5 AU from its parent star would lie at least this far off-axis for stars as distant as 10 pc, suggesting that stars as distant as 10 pc could be examined for planets fairly easily with large apodized apertures. Of course the reason that such detections have not in fact been possible to date is that diffracted light is not the whole story. In fact, the halo of scattered light due to wavefront irregularities (the ``error pattern'', to borrow a term from radio science) dominates the focal plane diffraction pattern beyond a few Airy rings from the central star [4,5], and so this topic is discussed in the next section.

3. Scattered Light

How debilitating is the scattered light halo? As is well known, the electric field distribution in the focal plane of an optical system is the Fourier transform of the electric field distribution in the final aperture plane. Another way to say this is that the diffraction pattern corresponds to the angular spectrum of the aperture plane electric field distribution. As a result, geometry determines the the diffraction pattern – the aperture's overall size, shape and illumination determine the ideal diffraction pattern, and the fine-scale structure determines the error pattern.

As an example, if the wavefront phase errors are described by a coherent wave at a single spatial frequency, the scattered power appears at two well-defined angular offsets from the center (similar to the ±1 orders in a sinusoidal grating). Wavefront errors with large lateral scales thus transform to the inner part of the diffraction pattern, and so it is errors at the larger spatial scales which need to be preferentially suppressed. In other words, because very fine lateral scale errors induce scattering very far from the axis, such errors are not of relevance for close-in planet searches. This eases the wavefront correction problem - what is required is control of wavefront phase (and amplitude) only over lateral spatial scales ranging from the full aperture diameter, D, to D/n, where $n\lambda/2D$ is the outermost angular radius to be corrected.

For simplicity, consider the case in which wavefront errors are dominated by phase errors (amplitude variations have similar implications). The Strehl ratio, S, defined as the ratio of the central intensities of the aberrated and ideal diffraction patterns, is given in the limit of small phase errors by

$$S = \exp(-\varphi_{rms}^2) = 1 - \varphi_{rms}^2, \qquad (1)$$

where φ_{rms} is the root-mean-square wavefront phase error. The fractional power scattered out of the main diffraction lobe is given by $1 - S = \varphi_{rms}^2$, and this power is distributed across the focal plane on scales corresponding to the power spectrum of the phase errors in the aperture plane. If there is a single dominant lateral error scale, d, then in a statistical sense the width of the scattered halo is approximately λ/d. The scattered power per pixel, i.e., per λ/D resolution element, relative to that in the central resolution element [4], is then of order $\varphi_{rms}^2(d/D)^2$. Assuming $n = D/d$ scattering centers across the aperture, the relative scattered power per pixel is approximately φ_{rms}^2/n^2. Thus, to reduce the scattered power level, a reduction is needed either in the magnitude of the phase deviations, in their lateral extents, or in both.

Assuming AO phase correction, a high number of actuators is then required to suppress the scattered light to the requisite level. To suppress off-axis light to a fraction, s, of the central pixel, we then have

$$\varphi_{rms} = ns^{1/2}. \qquad (2)$$

For a suppression ratio of 10^{-10}, we have $\varphi_{rms} < 10^{-3}$ radians for $n = 100$ actuators across the aperture diameter. The scattered light is suppressed out to n/2 resolution elements off axis, beyond which the AO system can no longer provide the needed correction [6]. The AO system must thus meet very demanding performance specifications, corresponding to a wavefront rms of about 0.1 nm ($\lambda/6000$) for a 100 x 100 system. Such highly accurate wavefront correction implies a level of sophistication beyond present capabilities.

We now address the relationship between the detectability of faint companions and the aperture diameter. Clearly, because the width of the diffraction-limited core of the PSF is proportional to D^{-1}, a planet at a given angular offset from the central star will appear further off-axis in angular resolution units for larger apertures. Thus, at a given field angle, the strength of the ideal stellar diffraction pattern fades as D increases. On the other hand, the scattered light reaching a given point in the focal plane originates in wavefront errors of a given spatial frequency, independent of the aperture diameter. As mentioned earlier, the relative scattered power per λ/D resolution element is approximately $\varphi_{rms}^2(d/D)^2$, which decreases as D^{-2} only if wavefront errors are corrected down to the same minimum spatial scale, independent of D. Since the number of actuators must then scale as $(D/d)^2$, larger telescopes require correspondingly more complex AO systems to achieve their theoretical gains.

4. Dimming the Central Star

It is however possible to reduce the brightness of the stellar image relative to its surroundings. Several techniques to accomplish this have been suggested, which generally fall into two categories: coronagraphy [8] and nulling interferometry [9]. In coronagraphy (Figure 3), the center of the star's diffraction pattern is modified by means of a focal plane mask, while in nulling interferometry (Figure 4), the

wavefronts arriving at two separate telescope apertures are superposed and subtracted in the pupil plane. In this sense the two techniques are complementary: nulling is basically an aperture plane technique, while coronagraphy can be described as essentially a focal plane technique.

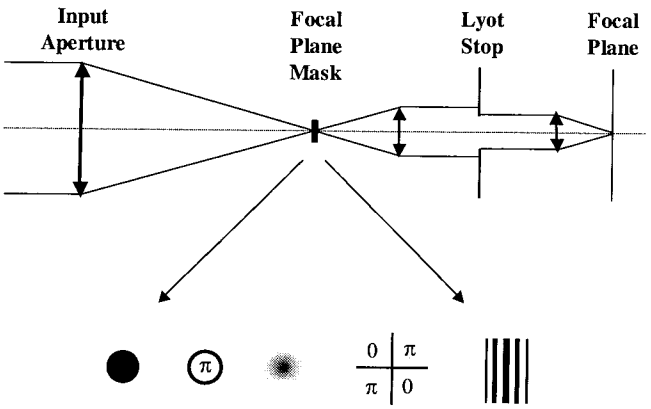

Figure 3. General coronagraph layout. A focal plane mask is located in the first focal plane, followed by an aperture (Lyot) stop at an image of the pupil. A number of proposed focal plane masks, explained in the text, are indicated along the bottom of the Figure.

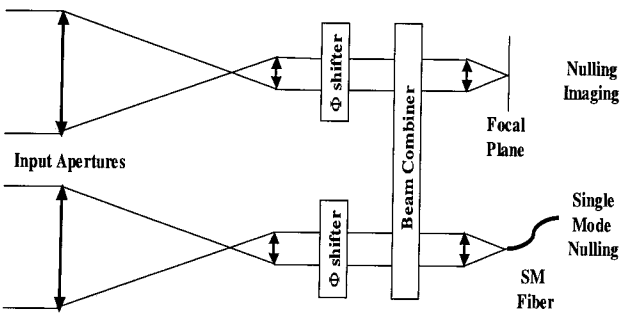

Figure 4. Layout of a general nulling interferometer. The phase shifters and the beam combiner operate in the aperture plane. The phase shifters effect a relative field reversal between the two incident beams. Spatial filtering can be applied in the focal plane, if desired, to achieve single-mode nulling.

However, this complementarity does not reflect a strict separation: coronagraphs also require modification of the aperture plane distribution via a Lyot (pupil) stop, while nullers can improve performance by means of spatial filtering in the final focal plane. Indeed, Figures 3 & 4 indicate that the operations are in some sense reversed in the two cases: nullers first modify (reverse and superpose) the aperture plane fields, after which spatial filtering can be applied in the focal plane, while coronagraphs first filter the focal plane PSF, and then modify (attenuate) the aperture plane distribution. Of course the focal plane filtering applied in the two cases may be vastly different: classical coronagraphs block out the central few Airy rings of the PSF, and so can detect objects only further off-axis than a few λ/D, while single-mode nulling (in which a single mode spatial filter is employed in the final focal plane) is only sensitive within the PSF's central lobe ($<\lambda/D$ off axis). Because of these different regimes of applicability, coronagraphy is naturally suited to searching for planets around the nearest stars, while nulling provides access to more distant stars and to closer in planets. Intermediate between these cases is nulling imaging, in which no spatial filtering is employed in a nuller's final focal plane, allowing access to the region further off-axis at the cost of reduced nulling and detection capabilities, and phase coronagraphy, in which the central obscuration is a phase mask instead of an opaque mask, allowing observational access closer to the axis than a classical opaque-mask coronagraph allows.

Of course, there are further differences between coronagraphy and nulling, one being that a coronagraph typically accepts the light from a single aperture telescope, while nulling interferometers typically operate by combining the beams from a number of individual apertures (be they separate telescope apertures or subapertures within a single larger telescope). To better understand the differences between these two techniques, we now describe each approach in a bit more detail.

5. CORONAGRAPHY

As described above, the basis of a coronagraph is the modification of the on-axis source PSF in the first focal plane. However, the Lyot stop is equally important, because it is in fact the Lyot stop which removes the high spatial frequency components of the energy in the aperture, i.e., the energy in the PSF's diffraction wings outside of the central obstruction which appears near the aperture rim in the pupil plane. Thus, while the focal plane mask removes most of the stellar flux (in the opaque mask case), it is the undersized Lyot stop which actually improves the contrast ratio for off-axis sources by suppressing the high spatial frequency PSF wings [10]. Of course, this improved contrast comes at the price of a lower throughput, due to the need to undersize the Lyot stop to remove the energy near the pupil rim.

For phase masks, the situation is actually reversed: because no energy is absorbed in the first focal plane, it is entirely up to the Lyot stop to remove the stellar energy. This is possible because phase masks result in a radical redistribution of the light in the pupil plane [11-13], essentially moving the light to the outside of the aperture image, where it can easily be blocked by an undersized stop. Thus, the focal-plane and pupil-plane masks are equally essential components of a coronagraph, and in general, the off-axis contrast ratio remains unaltered without the Lyot stop.

A number of different focal plane masks have been proposed, which fall into two categories: opaque masks and phase masks. There are several possible implementations in each of these categories, as indicated graphically along the bottom of Figure 3. From left to right, the cases depicted in this figure are: 1) a mask completely opaque over the central few Airy rings [8,10,14,15], 2) a mask which introduces a phase shift of π radians over the central Airy lobe [11,12], 3) a graded opacity mask, completely opaque in the center and increasing in transmission radially [16], 4) a 4-quadrant phase mask with a relative phase shift of π radians between the 1st and 3rd quadrants and the 2nd and 4th quadrants [13], and 5) a band-limited opaque mask which mimics a one-dimensional fringe pattern [17].

It is not the intent of this paper to fully delineate the specifics of each of these implementations. Rather, what is intended is a summary of their salient characteristics, in order to describe common operating principles. Luckily, there is a common framework for understanding all of these systems. First, all of the masks concentrate on modifying the PSF near the optical axis, i.e., either in removing the low spatial frequency components of the aperture plane distribution, or in modifying its phase. As such, the reimaged pupil plane contains energy largely near the aperture peripheries, which can then be removed by an undersized pupil (Lyot) stop.

The efficacy of the Lyot stop depends on its relative diameter and on the specific focal plane mask implementation. Of course there are two prices to be paid for the increased contrast available for off-axis sources: the region behind the central opaque obstruction is not available for observation in the opaque-mask case, and the throughput, even for the off-axis targets of interest, is necessarily reduced by the undersized pupil stop. The Lyot stop throughput decreases as the focal plane mask is made smaller, because the corresponding Lyot stop must then obscure a wider swath of the aperture periphery. However, these may not be high prices to pay if a large region near the optical axis is opened up to observation. In the phase mask case, the energy leaving the focal plane can in fact be redirected to the outside of the aperture image, so that the aperture plane field within the Lyot stop can in principle be reduced to zero for on-axis source in some implementations. Such phase coronagraphs thus resemble nullers in this regard.

In coronagraphs, the diffracted stellar energy is removed by absorption, either in the opaque focal plane mask itself, or by the opaque Lyot stop. On the other hand, scattered light is actually not diminished greatly, since this is already present in the first focal plane. Therefore, the same need for adaptive optics and phase correction exists with a coronagraph as for a simple imaging system, although bigger dividends are naturally expected with the coronagraph, because of the reduction in off-axis energy resulting from the attenuation of the aperture's high frequency components by the undersized Lyot stop.

Another important issue is the instantaneous bandwidth of the coronagraphic approach, i.e., the relative achromaticity of such approaches. There are two sources of chromatism. First, since all of the focal plane masks discussed are fixed in size, while the diffraction pattern scales with the wavelength, the effect of opaque and/or phase masks will typically depend on the size of the mask structures. However, there are exceptions: the 4-quadrant mask has no scale size, and so is achromatic from that point of view [13]. Likewise, the bandwidth limited mask can yield achromatic rejection because the energy transmitted by the fringe-like mask is so closely confined to the edges of the pupil that an undersized Lyot stop can in theory achieve 100% rejection of the residual flux [17]. However, manufacturing the needed focal plane fringe mask is very difficult, because of the need to transition from perfect rejection to perfect transparency across a small distance. A second bandwidth limitation for phase masks is that the induced phase shift needs to be an achromatic π radians. In practice this is difficult to achieve over broad bands, but one can conceive of narrow (few percent) to moderate (10-20%) bandwidth solutions. Narrow band coronagraphs should thus theoretically be able to achieve fairly high rejections, but bandwidths might easily be limited in practice by a number of manufacturing issues.

6. Nulling Interferometry

Nulling interferometry, on the other hand, is based on quite a different approach: the cancellation of the fields arriving at a number of separated apertures (either separate telescope apertures or subapertures within a single larger telescope) in the aperture plane. To effect the cancellation, the individual beams are matched in phase by an AO system and/or optical delay lines, and then superposed at an interferometric beam combiner (one or more beamsplitters). The optical system is then similar to a standard astronomical interferometer, except that the tolerances (e.g., on phase matching) are much tighter, and the high degree of symmetry required to effect achromatic cancellation calls for a more specialized (i.e. symmetric) beam combiner. One fundamentally different aspect of the nulling approach is that instead of using absorption to remove the stellar energy, in nulling the undesired starlight is instead simply reflected into unused beam-combiner output ports (the ``constructive''outputs).

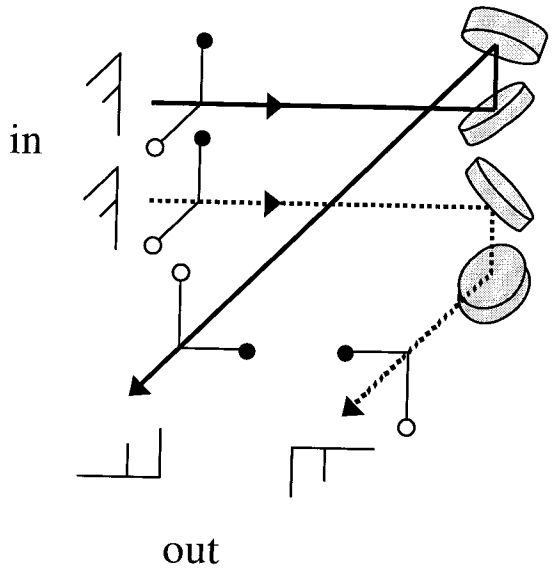

Figure 5. Illustration of the effect of a pair of reflective, mirror-symmetric, right-angle periscopes on two input beams. An achromatic relative field flip is introduced.

The basic goal of a nulling beam combiner is the superposition of a pair of beams with reversed electric fields. For deep nulling, the field reversal and beam combination stages need to meet stringent requirements on achromaticity, polarization-independence, field rotation, intensity matching, and pathlength (phase) matching [18]. A completely symmetric nulling beam combiner design has recently been described [19], which first introduces a relative field flip between the two incident beams, and then superposes the beams at a perfectly-symmetric constructive beam combiner. These two components of the fully symmetric nulling beam combiner are shown in Figures 5 and 6. Figure 5 shows a means of introducing a perfectly achromatic field reversal by means of a geometric field flip [19]. Alternatively, a relatively achromatic phase delay of π radians can also be introduced by first sending the two beams through two sets of dielectric plates of slightly differing thicknesses [20]. Figure 6 shows the beam combiner itself, which can work with either type of phase shifter, and which achieves complete symmetry by passing the beams to be combined through a pair of reciprocal beamsplitters (in which whatever happens to one beam at one beamsplitter happens to the other beam at the other splitter). Given that this nulling beamcombiner design relies entirely on flat optics, is only slightly more complicated than standard beamcombiners already employed in astronomical interferometers (except for finer tolerances), and places no great demands on the beamsplitter design (other than on the matching of the beamsplitter pair) this nulling approach is fairly straightforward to implement, and so the issue of nulling beamcombiner design is no longer an obstacle.

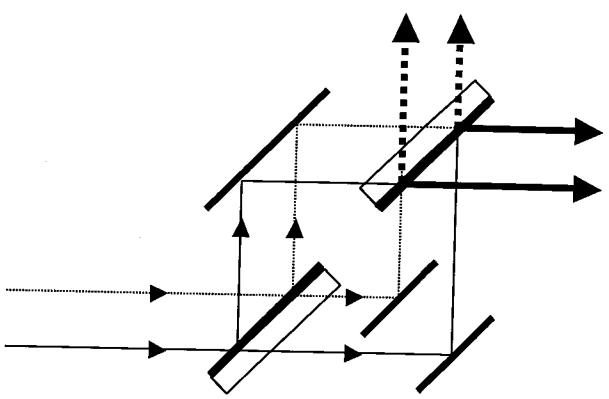

Figure 6. Modified Mach-Zehnder beam combiner. The two beams to be combined enter at the lower left, and the symmetric (constructive) outputs exit toward the right. The three black solid lines represent mirrors, and the two white rectangles are beamsplitter substrates. The beamsplitter coatings are represented by the dark surfaces on the substrates. Although this layout in fact represents a constructive beam combiner, the introduction of a prior field flip converts it to a nulling beam combiner.

In fact, deep and stable experimental nulls have already been demonstrated in the laboratory (Figure 7) at optical wavelengths [21,22], albeit with a different nulling approach. Almost all of the essential physics inherent in nulling has therefore been demonstrated in the laboratory. However, a great deal of work remains in terms of translating these results into the infrared, and in pushing the technology to reach the deep, broadband, dual-polarization nulls demanded by the TPF performance goals.

Is then nulling the obvious approach to use for high contrast near-neighbor detection? Unfortunately, matters are not quite so simple, for several reasons. First, the wavelength range for which nulling is usually foreseen is in the mid-infrared (roughly 5-20 μm), because of the reduced planet/star contrast ratio in the thermal infrared (10^{-6} in the thermal infrared vs. 10^{-10} at optical wavelengths). However, to achieve an angular resolution comparable to a large-aperture visible-wavelength telescope, such long wavelengths require proportionally longer baselines (several tens of meters). This brings with it the need for ether a large, precise structure, or separated spacecraft and formation flying. Second, infrared nulling brings with it the need to cool both the entire optical train and the detectors to cryogenic temperatures, a major cost driver if a number of separated spacecraft are involved.

A saving grace is that he beams to be nutlled must be phased to an accuracy [18] of

$$\varphi = 2N^{1/2}, \tag{3}$$

or about 3 nm for mid-infrared nulls, N, of 10^{-6}. This is not nearly as exacting as the phasing requirements a visible wavelengths (Section 3).

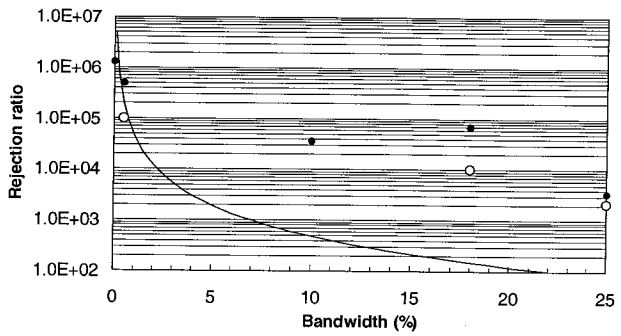

Figure 7. Null depths achieved experimentally with the JPL rotational shearing interferometer vs. source bandwidth. The filled points give the best transient rejection ratios measured, while the open circles give the best stabilized null levels achieved. The curve gives the best rejection level achievable with the standard chromatic dark fringe in a constructive interferometer. Several points are well above this level, indicating that achromatic nulling has been achieved.

On the other hand, what about nulling at optical wavelengths? Here the nulls would have to be much deeper, 10^{-9} - 10^{-10} vs. the easier MIR requirement of 10^{-6}, and so the rejection and phase control needed in an optical nuller is comparable to that in an optical coronagraph. A second problem is that since such deep nulls can only realistically be attained by means of single-mode nulling, the resultant single-mode field of view at optical wavelengths would be tiny. To overcome these difficulties, subaperture nulling has been proposed [23]. In this approach, a typical optical telescope is divided into a number of smaller subapertures, which then have larger single-mode diffraction beamwidths (and hence fields of view). To phase each of the sub-nullers individually, a deformable mirror (DM) would be needed, as in AO, with individual DM elements matched to individual subnullers. In essence, the wavefront would be corrected to allow for the nulling of each subaperture involved. The individually nulled subapertures would then be recombined at a final large imaging lens. Because the nulled image consists of n^2 pixels, where n^2 is the number of subapertures, the final rejection ratio at an off axis pixel is then

$$\rho = N(1-S)/n^2, \tag{4}$$

where N is the null depth (rejection) achieved in the nulling stage, S is the Strehl ratio achieved in the imaging stage, and n^2 is the number of pixels in the final image (equal to the number of DM elements). Using equation 3, the phase accuracy needed is then

$$\varphi = 2n\rho^{1/2}/(1-S)^{1/2}, \tag{5}$$

or, assuming $S \approx 0.8$,

$$\varphi \approx 4n\rho^{1/2}. \tag{6}$$

Except for the coefficient, equation 6 is identical to equation 2, implying similar phasing and actuator-spacing constraints for optical subaperture nulling and coronagraphy.

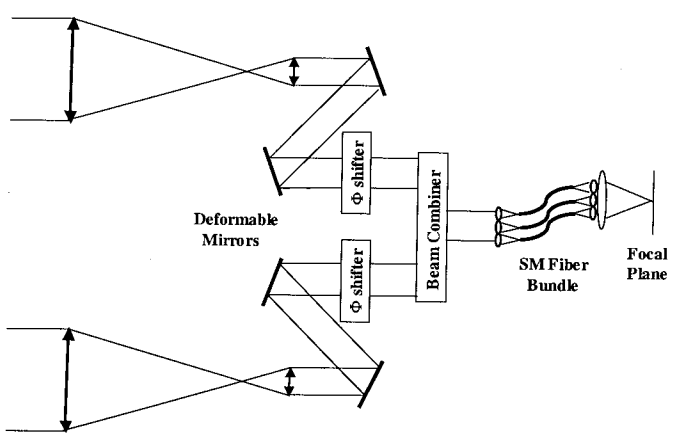

Figure 8. Layout of a subaperture nuller. The beam aperture is divided into a number of subapertures, each one of which is nulled and spatially filtered individually. The subapertures are defined by the lenslet array elements, which map back to deformable mirror elements that provide the capability to phase the individual subapertures.

7. Comparison Summary

To detect planets as faint as the Earth near bright stars, the light from stars will need to be dimmed in comparison to that from their surroundings using either coronagraphy or nulling, at least until techniques for high-accuracy, large-aperture, space-based telescopes are developed. Coronagraphy greatly eases the requirements on the aperture size, and it may in fact be possible to detect certain classes of planet with an optical coronagraph operating on space-based telescopes of a few meter diameter. However, while new types of coronagraph proposed recently may provide further theoretical advantages, scattering from small scale surface irregularities limits the ability of any coronagraph to reach the high contrast ratios desired in practise without highly accurate fine-scale wavefront correction. On the other hand, infrared nulling interferometry, in which the light from two or more apertures is combined destructively in order to subtract the starlight to high accuracy, reduces the necessary phase control level by about two orders of magnitude. However, infrared nulling requires multiple separated cryogenic telescopes, as well as cryogenic optical trains, and so brings its own difficulties. Optical subaperture nullers, meanwhile, have very similar phase requirements to optical coronagraphs, and the apertures are of similar sizes, so that these two approaches are of comparable difficulty.

Thus, both coronagraphs and nulling interferometers must face and solve nearly identical phasing issues. On the other hand, in the thermal infrared, this requirement is substantially reduced, both because of the longer wavelengths, and the reduced contrast ratio. This however comes at the cost of much longer baselines, a higher background noise level, and the need to operate the entire optical train cryogenically. Thus, none of these techniques is as yet particularly easy, and all are expensive. Further technology development and a precursor mission to TPF are thus highly desirable, the former to identify favored approaches as soon as possible, and the latter to search for the brighter extra-solar planets on a more rapid timescales.

Not covered is a discussion of signal to noise.

The research described herein was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

[1] http://exoplanets.org, http://www.obspm.fr/planets

[2] P. Nisenson & C. Papaliolios, "Detection of Earth-like Planets Using Apodized Telescopes," *Astroph. J. Lett.* **548**, *L201-L205*, 2001.

[3] D. Spergel, 2002, preprint.

[4] J. Ruze, "Antenna Tolerance Theory – A Review," *Proc. IEEE*, **54**, *633-640*, 1966.

[5] J.B. Breckinridge, T.G. Kuper & R.V. Shack, "Space Telescope Low-scattered-light Camera: A Model," *Opt. Eng.* **23**, *816-820*, 1984.

[6] F. Malbet, J.W. Yu & M. Shao, "High-Dynamic-Range Imaging Using a Deformable Mirror for Space Coronagraphy," *Publ. Astron. Soc. Pac.* **107**, 386-398, 1995.

[7] J.T. Trauger, priv. comm., 2001.

[8] B. Lyot, "A Study of the Solar Corona and Prominences without Eclipses," *Mon. Not.R. Astron. Soc.*, **99**, *580*, 1939.

[9] E. Serabyn, "Nulling Interferometry and Planet Detection," in *Principles of Long Baseline Stellar Interferomety*, ed. P.R. Lawson, JPL Pub. 00-009, *275-292*, 2000.

[10] A. Sivaramakrishnan, C.D. Koresko, R.B. Makidon, T. Berkefeld, & M.J. Kuchner, "Ground-based Coronagraphy with High-order Adaptive Optics," *Astrophys. J.*, **552**, *397-408*, 2001.

[11] F. Roddier & C. Roddier, "Stellar Coronagraph with Phase Mask," *Publ. Astron. Soc. Pac.* **109**, *815-820*, 1997.

[12] O. Guyon et al. "The Nulling Stellar Coronagraph: Laboratory Tests and Performance Evaluation," *Publ. Astron. Soc. Pac.* **111**, *1321-1330*, 1999.

[13] D. Rouan, P. Riaud, A. Boccaletti, Y. Clenet & A. Labeyrie, "The Four-Quadrant Phase-Mask Coronagraph. I. Principle," *Publ. Astron. Soc. Pac.* **112**, *1479-1486*, 2000.

[14] D.A. Golimowski, M. Clampin, S.T. Durrance & R.H. Barkhouser, "High-resolution Ground-based Coronagraphy Using Image-Motion Compensation," Appl. Opt. **31**, *4405-4416*, 1992.

[15] F. Malbet, "High Angular Resolution Coronagraphy for Adaptive Optics," *Astron. Astrophys. Suppl.* **115**, *161-174*, 1996.

[16] J.T. Trauger, "High Contrast Coronagraphic Imaging with NGST," in *NGST Science & Technology Exposition*, Astron. Soc. Pac. Conf. Ser. Vol. **207**, *165-178*, 2000.

[17] M.J. Kuchner & W.A. Traub, "A Coronagraph with a Bandwidth-Limited Mask for Finding Terrestrial Planets," Astrophys. J. in press.

[18] E. Serabyn, "Nulling Interferometry: Symmetry Requirements and Experimental Results," in *Interferometry in Optical Astronomy*, Proc. SPIE **4006**, *328-339*, 2000.

[19] E. Serabyn & M. M. Colavita, "Fully Symmetric Nulling Beam Combiners," Appl. Opt. **40**, *1668-1671*, 2001.

[20] R.M. Morgan, J. Burge, & N. Woolf, "Nulling interferometric beam combiner using dielectric plates: experimental results in the visible broadband," in *Interferometry in Optical Astronomy*, Proc. SPIE **4006**, *340-348*, 2000.

[21] K. Wallace, G. Hardy & E. Serabyn, "Deep and Stable Interferometric Nulling of Broadband Light with Implications for Observing Planets around Nearby Stars," Nature **406**, *700-702*, 2000.

[22] E. Serabyn, S.R. Martin & G.J. Hardy, "Progress Toward Space-based Nulling Interferometry: Comparison of Null Stabilization Approaches," in Proc. 2001 IEEE Aerospace Conference, 2001.

[23] M. Shao, "Hubble Extra Solar Planet Interferometer," in *Space Astronomical Telescopes and Instruments*, Proc. SPIE **1494**, *347-356*, 1991.

Gene Serabyn is a senior research scientist at JPL.